ASSESSMENT OF DIFFERENT NUMERICAL SOLUTION STRATEGIES FOR GRAVITY FIELD RECOVERY

W.-D. Schuh[1], R. Pail[2], G. Plank[2]

[1] ***Institute of Theoretical Geodesy, University of Bonn,***
Nussallee 17, 53115 Bonn, Germany.
Email : schuh@theor.geod.uni–bonn.de

[2] ***Institute of Theoretical Geodesy, Graz University of Technology,***
Steyrergasse 30, 8010 Graz, Austria.
Email : pail@geomatics.tu–graz.ac.at / plank@geomatics.tu–graz.ac.at

INTRODUCTION

The processing of the GOCE observations will be a laborious task. The mission concept, i.e. the combination of high-low SST and satellite gravity gradiometry, will lead to several 100 millions of data holding – among other contributions – the Earth's gravity field information. The retrieval of this field, parameterized by about 100.000 unknowns, requires sophisticated tools of numerical mathematics.

In the last decade several approaches have been developed to solve this huge system of equations. While the *direct method* directly processes the gravity field observations in space domain, the *time-wise approach* considers the observations along a satellite track as a time series. Although the observation equations are not identical for the two methods, both processing strategies should lead to nearly identical results in terms of the Level 2 products.

Besides, one crucial aspect of GOCE data processing is the adequate treatment of the instrument noise. Intelligent filter strategies have to be applied to cope with the colored noise behaviour of the satellite's gravity sensors.

In the present paper an assessment of the different methods will be performed. We will provide both a theoretical and a numerical comparison of the two main processing strategies by means of a closed loop simulation, demonstrating the advantages, the drawbacks and the practical feasibility, especially emphasizing and discussing the problem of a correct treatment of the measuring noise.

THEORY

The gravitational potential of the Earth is parameterized by a harmonic series expansion in spherical coordinates (r, θ, λ)

$$V(r,\theta,\lambda) = \frac{GM}{R}\sum_{l=0}^{\infty}\left(\frac{R}{r}\right)^{l+1}\sum_{m=0}^{l}\bar{P}_{lm}(\cos\theta)\,[\bar{C}_{lm}\cos m\lambda + \bar{S}_{lm}\sin m\lambda]. \qquad (1)$$

G, M and R are the gravitational constant, the Earth's mass and the Earth's reference radius, respectively, while $\bar{P}_{lm}$ denotes the fully normalized Legendre polynomials depending on the harmonic degree l and order m. In practice, this series expansion has to be cut off at an upper maximum degree l_{max}, i.e. the upper limit ∞ of the index l is replaced by a finite number l_{max}, defining the resolution of the parameterization.

The goal is now to determine the corresponding harmonic coefficients $\bar{C}_{lm}$ and $\bar{S}_{lm}$ as accurately as possible.

The dedicated gravity gradiometry mission GOCE will deliver essentially two types of measurements: GPS satellite-to-satellite tracking (SST) data reflecting orbit perturbations and representing first order derivatives of the gravitational potential, as well as gravity gradients (SGG) measured by the 3-axis gravity gradiometer representing second order derivatives. In the present paper we will concentrate on the latter measurement type.

Gathering all the second order derivatives with respect to the main directions (radial, along track, across track) in an observation vector ℓ and the corresponding error covariances in the matrix Σ, using a standard Gauss-Markov model and applying the best linear uniformly unbiased estimation with respect to the Σ^{-1}-norm leads

to the normal equation system

$$(A^T\Sigma^{-1}A)\,\hat{x} = A^T\Sigma^{-1}\ell \tag{2}$$

for the parameter vector $\hat{x} = \{\bar{C}_{lm}; \bar{S}_{lm}\}$, where the design matrix A represents the linear (or linearized) relation between observations ℓ and the parameters x and thus contains the base functions of the series expansion (1). This approach is in principle applied by all *direct methods* to solve for the coefficients.

Due to the fact that measurements taken along a satellite's track are not homogeneously distributed with respect to the spatial domain of the sphere, the orthogonality properties of the base functions are lost, and thus the normal equation matrix $A^T\Sigma^{-1}A$ is a dense matrix. However, if the gravity gradients are uninterruptedly measured with constant step size along a circular repeat orbit, it can be shown that all coefficients of different orders m as well as *cos-* and *sin*-related coefficients become uncorrelated, leading to a block-diagonal structure of the equation system (cf. [4]).

In reality, these simplifying conditions will not be strictly fulfilled, and consequently the normal equation system will deviate from the block-diagonal structure. However, due to the large number of unknown parameters and thus also the huge dimension of the normal equation system $A^T\Sigma^{-1}A$ (about 90000 unknowns for a maximum resolution of $l_{max} = 300$, corresponding to 70 km half wavelength), the system can not be directly inverted.

Therefore, for the processing of real mission data the *direct method* was implemented in terms of the `pcgma` (**P**reconditioned **C**onjugate **G**radient **M**ultiple **A**djustment) algorithm (cf. [5]), using a block-diagonal preconditioner $(A^T\Sigma^{-1}A)_o$ as a first guess (initialization step) and applying iteratively a conjugate gradient method to successively improve the parameter estimates.

While in the *direct method* the observations are regarded as functions of the geographical location (r, θ, λ), they can also be considered as a periodic time series for one repeat period (cf. [3], [4]). Assuming a circular orbit, (1) can be rewritten as

$$V = \frac{GM}{R}\sum_l \left(\frac{R}{r}\right)^{l+1} \sum_m \sum_k \bar{F}^k_{lm}(i)[\alpha_{lm}\cos\psi_{km}(t) + \beta_{lm}\sin\psi_{km}(t)] \tag{3}$$

with $\bar{F}^k_{lm}(i)$ denoting the inclination function which depends on the orbit inclination i, and the coefficients α_{lm}, β_{lm} related to the harmonic coefficients $\bar{C}_{lm}$, $\bar{S}_{lm}$. Abbreviating

$$\left\{ \begin{array}{c} A_{km} \\ B_{km} \end{array} \right\} = \frac{GM}{R}\sum_l \left(\frac{R}{r}\right)^{l+1} \bar{F}^k_{lm}(i) \left\{ \begin{array}{c} \alpha_{lm} \\ \beta_{lm} \end{array} \right\} \quad , \tag{4}$$

(3) is a Fourier series

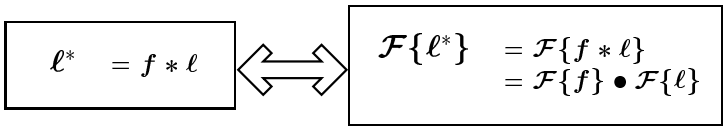
$$V = \sum_m \sum_k [A_{km}\cos\psi_{km}(t) + B_{km}\sin\psi_{km}(t)] \tag{5}$$

where $\psi_{km}(t)$ is related to two fundamental frequencies ω_o (orbit revolution) and ω_e (Earth's rotation).

In this *time-wise approach* (also frequently denoted as *semi-analytic approach*) in the first step the Fourier coefficients ('lumped coefficients') A_{km}, B_{km} are computed by FFT methods, and in the second step the harmonic coefficients $\bar{C}_{lm}$, $\bar{S}_{lm}$ are estimated order by order, assuming strict block-diagonal structure of the system. The deviations from this property are incorporated by means of an iterative procedure.

Summarizing, while the *direct method* works in the space domain, the *semi-analytic approach* switches to the time domain and the corresponding frequency domain. This major difference becomes evident including the noise behaviour of the gravity sensor. The gradiometer has a selected measurement bandwidth, where it measures the gravity gradients very accurately, while the performance beyond this spectral window is substantially degraded. Naturally this colored noise behaviour has to be implemented and will be reflected by the observation covariances Σ.

Fig. 1 summarizes the operations which have to be performed in the space domain (left) and the frequency domain (right), respectively. Introducing the sensor's noise characteristics can be considered as a filter which has to be applied to both the observations ℓ and to the normal equations. Reference [6] provides a method for the implementation of recursive digital filters, essentially representing a convolution of the observations ℓ with a filter impulse response f, leading to the filter result $\ell^* = f * \ell$.



time domain	frequency domain
ℓ ... signal	$\mathcal{F}\{\ell\}$ ... signal
ℓ^* ... filtered signal	$\mathcal{F}\{\ell^*\}$ ... filtered signal
f ... impulse response	$\mathcal{F}\{f\}$ ... frequency response

Fig. 1. Filtered random process; time domain vs. frequency domain.

On the other hand, due to the fact that the 'lumped coefficients' A_{km}, B_{km} – representing the observations in frequency domain $\mathcal{F}\{\ell\}$ – are directly related to a frequency via ψ_{km}, in the time-wise approach in frequency domain this filter operation is a simple multiplication.

SIMULATION

In order to demonstrate the performance of the different approaches, we generate test data sets with the intention to simulate the instrument noise characteristics as realistically as possible. In this section we describe two different strategies to generate random variables with a specific spectral behaviour. On the one hand a harmonic random process simulates a periodic behaviour with a discrete spectrum, while on the other hand a general linear process generates a non-periodic noise with a purely continuous spectrum. In reality a combination of both effects, periodic and non-periodic, will degrade the measurement signal.

A time-discrete random process may be considered as an indexed sequence of random variables

$$\ldots,\, x(-2),\, x(-1),\, x(0),\, x(1),\, x(2),\, \ldots \tag{6}$$

where the statistics of each random variable in the sequence $\{x(n)\}$ is defined by a probability distribution. This ensemble of discrete time signals generates a sequence of *mean values* as a function of time n

$$m_x(n) = \mathsf{E}\left\{x(n)\right\} \tag{7}$$

and an ensemble of *variances*

$$\sigma^2(n) = \mathsf{E}\left\{\left[x(n) - m_x(n)\right]^2\right\} \tag{8}$$

of the process. Two other statistic moments are the *autocovariance*

$$c_x(k,l) = \mathsf{E}\left\{\left[x(k) - m_x(k)\right]\,\left[x(l) - m_x(l)\right]\right\} \tag{9}$$

and the *autocorrelation*

$$r_x(k,l) = \mathsf{E}\left\{x(k)\,x(l)\right\}\quad . \tag{10}$$

For our approach, we consider in particular *wide sense stationary* (WSS) processes. Hence, the following three conditions must be satisfied:

- the mean of the process is independent of time, $m_x(n) = m_x$,
- the autocorrelation $r_x(k,l)$ depends only on the difference $k - l$, and
- the variance of the process is finite, $c_x(0) = \sigma_x^2 < \infty$.

The autocorrelation sequence of a WSS process provides a description of the second-order moment of the process in time domain. If we compute its discrete Fourier transform

$$P_x\left(e^{j\omega}\right) = \sum_{k=-\infty}^{\infty} r_x(k)\; e^{-jk\omega} \quad , \tag{11}$$

we obtain the *power spectral density* (power spectrum) of the process, which characterizes the second order moment of the process in the frequency domain. The *total power* of the process is proportional to the area under the power spectral density curve

$$\mathsf{E}\left\{|x(n)|^2\right\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_x\left(e^{j\omega}\right)\, d\omega \quad . \tag{12}$$

Standard discrete parameter models can be divided into two quite different groups: the general linear processes and the harmonic processes. The *harmonic process* is defined by

$$x(n) = \sum_{i=1}^{I} A_i \cos(n\omega_i + \phi_i) \tag{13}$$

where $\{A_i\}, \{\omega_i\}, (i = 1, \ldots, I)$ are constants and the phases $\{\phi_i\}, (i = 1, \ldots, I)$ are independent random variables, uniformly distributed between $[-\pi\ ,\ \pi]$. This process is characterized by a zero mean

$$m_x = \mathsf{E}\left\{x(n)\right\} = 0 \tag{14}$$

and the autocorrelation (= autocovariance) sequence

$$r_x(k,l) = r_x(k-l) = \frac{1}{2}\sum_{i=1}^{I} A_i^2 \cos\left((k-l)\omega_i\right) \tag{15}$$

which depends only on the time difference $k - l$. Thus, the harmonic process is a WSS process. The power spectral density can be determined by

$$P_x\left(e^{j\omega}\right) = \frac{1}{2}\pi \sum_{i=1}^{I} A_i^2 \left[\delta(\omega - \omega_i) + \delta(\omega + \omega_i)\right] \tag{16}$$

where $\delta(\omega - \omega_i)$ is used to denote an impulse at frequency $\omega = \omega_i$, and $\delta(\omega + \omega_i)$ is an impulse at $-\omega_i$. If, in addition, the amplitudes $\{A_i\}$ are also uncorrelated random variables, then the autocorrelation sequence is

$$r_x(k,l) = r_x(k-l) = \frac{1}{2}\sum_{i=1}^{I} \mathsf{E}\{A_i^2\} \cos\left((k-l)\omega_i\right) \quad , \tag{17}$$

and the power spectral density can be computed by

$$P_x\left(e^{j\omega}\right) = \frac{1}{2}\pi \sum_{i=1}^{I} \mathsf{E}\{A_i^2\} \left[\delta(\omega - \omega_i) + \delta(\omega + \omega_i)\right] \quad , \tag{18}$$

again defining a WSS process.

In contrast to the harmonic process a *general linear process* is given by the convolution

$$y(n) = x(n) * h(n) = \sum_{k=-\infty}^{\infty} h(k)\ x(n-k) \tag{19}$$

of a WSS random sequence $x(n)$ by a linear shift-invariant filter defined by the unit sample response $h(n)$. The mean of this process

$$m_y(n) = E\{y(n)\} = E\left\{\sum_{k=-\infty}^{\infty} h(k)\ x(n-k)\right\} = m_x \sum_{k=-\infty}^{\infty} h(k) = m_y \tag{20}$$

is independent of the time n, and thus the autocorrelation sequence

$$r_y(k) = r_x(k) * h(k) * h(-k) \tag{21}$$

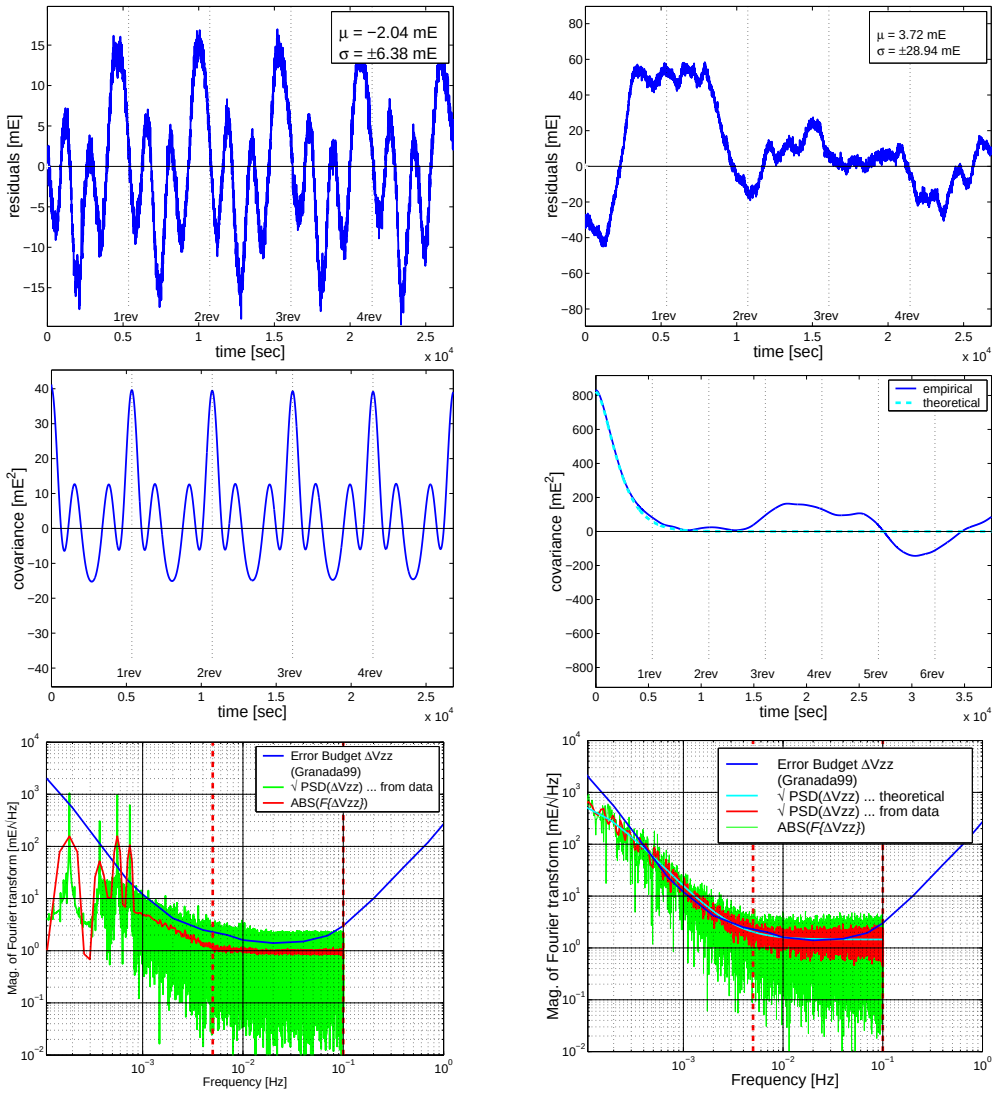


Fig. 2. Colored noise generated by a harmonic random process (left) and a filtered random process computed by an autoregressive moving average process.
Left column: *Delft-data-set* ΔV_{zz}; right column: *ARMA-data-set* ΔV_{zz};
(top) signal; (center) empirical covariance function; (bottom) periodogramm and square root of the power spectrum density.

is also shift-invariant. Therefore the power spectral density of the filtered process is given by

$$P_y\left(e^{j\omega}\right) = P_x\left(e^{j\omega}\right)\left|H\left(e^{j\omega}\right)\right|^2 \tag{22}$$

where $H\left(e^{j\omega}\right)$ denotes the frequency response of the filter. In terms of z-transforms this is written by

$$P_y(z) = P_x(z)\,H(z)\,H^*(1/z^*) \quad . \tag{23}$$

If we use a causal linear shift-invariant filter of the form

$$y(n) + \sum_{l=1}^{p} a_l\, y(n-l) = \sum_{l=0}^{q} b_l\, x(n-l) \tag{24}$$

(autoregressive moving average, ARMA filter), the frequency response function is directly given by

$$H(z) = \frac{\sum_{l=0}^{q} b_l \, z^{-l}}{1 + \sum_{l=1}^{p} a_l \, z^{-l}} \tag{25}$$

Thus, if we use a white noise process as input process, the frequency response function of the filter determines the spectral behaviour of the output process $y(n)$.

Concerning the satellite mission GOCE the error budget of the gradiometer output was studied (cf. [1]), and the spectral behaviour of the measurements is given in terms of the power spectral density curves (cf. Fig. 2).

Based on this spectral information, test data sets are generated to study the performance of different solution strategies. In the frame of the ESA study [2] the so-called *Delft-data-set* was generated by SID/DEOS (cf. [2], pp. 63-66 and pp. 138-141). The simulated orbit is a sun-synchronous, 29-day repeat orbit, with an inclination of $i = 96.6°$ and an altitude of approximately 250 km. The simulated measurements (3 main diagonal elements of the gradient tensor) are regularly distributed along this nearly circular orbit with a sampling interval of 5 sec. The colored noise is generated by a random harmonic process. As an example, the left column of Fig. 2 shows the statistical behaviour of the radial component ΔV_{zz} of this data set, which demonstrates the specific property of a harmonic process, i.e. a periodic behaviour of the signal, and a non-decreasing periodic covariance function. The Fourier analysis of the data reflects also the discrete characteristic of the the power spectra. On the basis of the same orbit data a second data set is generated, where the colored noise is simulated by an ARMA filter process. We use the filter coefficients described in detail in [2], pp. 155-157 and change the numerator and denominator of the frequency response function, which means an interchange of the zeros with the poles. In order to stabilize the behaviour, we additionally introduce a stop band filter for zero and extremely low frequencies. The right column of Fig. 2 shows the statistical behaviour of this data set, which is denoted in this article as *ARMA-data-set*.

RESULTS

This closed loop simulation is based on more than 1.5 million observations (29 days repeat orbit, 5 s sampling) of the three main diagonal elements of the gradient tensor containing information about the Earth's gravity field up to degree and order 180 (based on the Earth model OSU91a), superposed by a noise model generated by a harmonic process (*Delft-data-set*) and by an filter process (*ARMA-data-set*) as described in the previous section. Within all our simulations we can not detect a significant difference in the solution behaviour using the different noise models. Consequently, in the following documentation we concentrate on the *ARMA-data-set*.

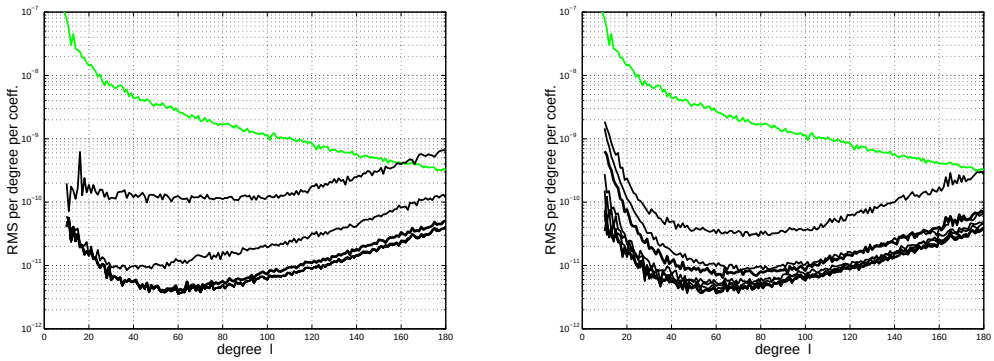


Fig. 3. Convergence behaviour: *Semi-analytic approach* (left) versus *direct method* (right). Gray line: Spectrum OSU91a, black lines: Differences between exact and iterative solutions σ_l as defined by (26).

The *direct method* as well as the *semi-analytic approach* were applied to estimate the set of 32757 coefficients describing the Earth's gravity field up to a spatial resolution of about 110 km half wavelength. Due to the fact that the input data set is based on a realistic, perturbed orbit, both methods have to apply an iterative solution strategy, computing 15 iterations for each approach.

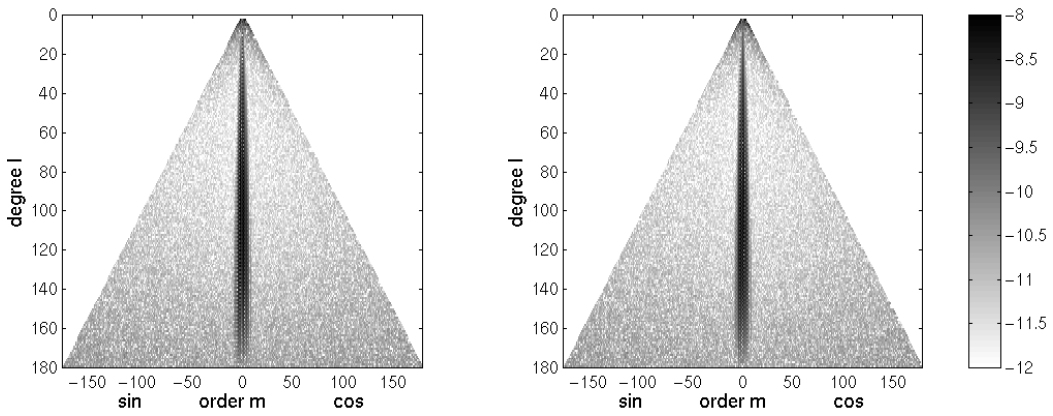


Fig. 4. Spectral triangles: *Semi-analytic approach* (left) versus *direct method* (right). Deviations of coefficients after the 15. iteration from exact solution. Plotted is $log_{10}(|\ldots|)$.

Fig. 3 shows the deviations of the estimated coefficients from the initial 'exact' OSU91a-model in terms of a degree rms

$$\sigma_l = \sqrt{\frac{1}{2\,l+1}\sum_{m=0}^{l}[(\bar{C}_{lm}^{(est)} - \bar{C}_{lm}^{(OSU)})^2 + (\bar{S}_{lm}^{(est)} - \bar{S}_{lm}^{(OSU)})^2]} \tag{26}$$

Except of a slightly differing convergence behaviour, it can be stated that both methods reach the final solution after four or five iterations, and that both results are nearly identical. This statement is underlined by Fig. 4, which illustrates the absolute deviations of the coefficient estimates from the truth. Due to the polar gap problem (no observations in the polar regions because of the sun-synchronous orbit design), the zonal and near-zonal coefficients ($m < 10$) are resolved worst.

The quality of the solutions can also be compared in terms of second level products such as geoid heights, thus closing the loop of our simulation strategy. Fig. 5 shows the absolute geoid height deviations in $[m]$ of the *semi-analytic approach* (left) and the *direct method* solution (right) after the 15. iteration. In the first row we compare the global solution. Although the actual error structures are different, they have the same order of magnitude for both approaches.

In the second row we illustrate a regional solution for the equiangular test area 10° W to 55° E in longitude and 10° N to 65° N in latitude. The global trend was removed subtracting a trend field which was generated applying a simple 30° × 30° moving average filter. Reducing the long-wavelength errors, the two regional approaches show the capability to reconstruct the high frequency behaviour.

Table 1. ARMA-Data Set ($l_{max} = 180$); reconstruction of second level information: Numerical comparison *semi-analytic approach* versus *direct method*. Beside the global behaviour (incl. and excl. polar regions) the local approaches shows the capability to reconstruct the high frequency behaviour. For the local representation the data are reduced by a moving average (30° × 30°).

5 iterations	global		$-83° \le \phi \le 83°$		local			
	min	max	min	max	min	max	mean	σ
semi-analytic approach	-64	67	-0.283	0.228	-0.088	0.086	-0.001	0.025
direct approach	-10825	12616	-5.917	2.722	-0.106	0.147	0.008	0.037
15 iterations	global		$-83° \le \phi \le 83°$		local			
	min	max	min	max	min	max	mean	σ
semi-analytic approach	-63	67	-0.285	0.228	-0.088	0.086	-0.001	0.025
direct approach	-69	87	-0.289	0.389	-0.082	0.079	-0.003	0.024

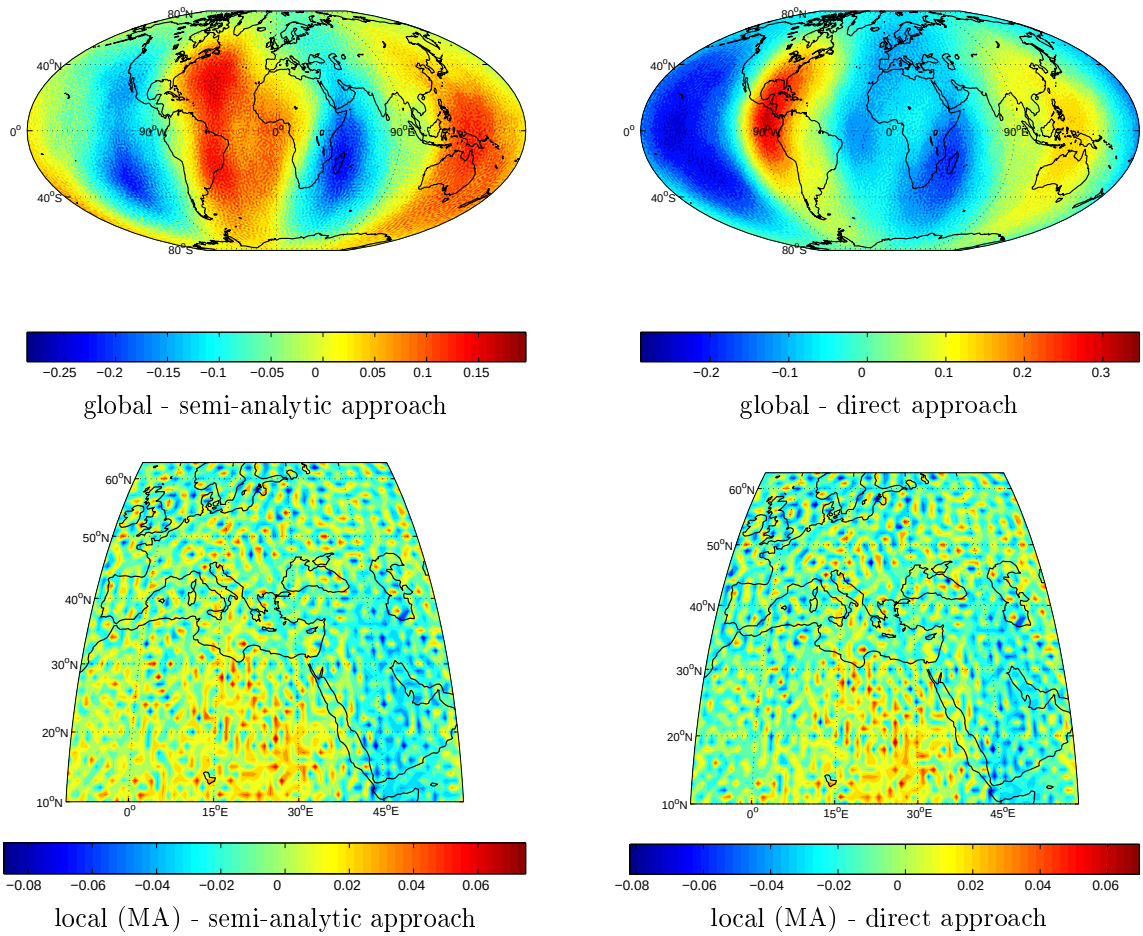


Fig. 5. ARMA-Data Set ($l_{max} = 180$); reconstruction of second level information: Comparison *semi-analytic approach* versus *direct method* (both after 15 iterations). Beside the global behaviour the local approaches show the capability to reconstruct the high frequency behaviour. For the local representation the data are reduced by a moving average ($30° \times 30°$).

Table 1 summarizes some statistical parameters of the geoid height difference fields for the two approaches after the 5. and 15. iteration, respectively.

The only significant difference between the two approaches occurs in the 5. iteration for the global solution, which can be explained by a worse convergence behaviour of the `pcgma` with respect to the long wavelengths and low order coefficients (cf. Fig. 3).

DISCUSSION

The results discussed in the previous section demonstrate that two completely different and independent solution strategies, i.e. the *direct method* implemented in terms of the `pcgma` algorithm as well as the *time-wise (semi-analytic) approach*, lead to essentially identical results. The major conceptual difference lies in the treatment of the system in the space versus time/frequency domain, respectively, which is directly linked to a different incorporation of the colored measuring noise and the implementation of filters. While the *direct method* computes a convolution of the filter function in space domain, the *semi-analytic approach* provides a more direct access to the frequency characteristics of the gradiometer, which can be mathematically expressed by a simple multiplication in frequency domain. For the point of view of the computation effort, the semi-analytic approach

is much cheaper, resulting in a faster recovery of the coefficient estimates but with the same order of accuracy. Test computations on the same machine (DEC Alpha, Workstation XP1000, 1 GB RAM, Alpha 21246 CPU) demonstrate that at the present state of implementation of the two processing strategies the *semi-analytic approach* outperforms the `pcgma` by about a factor of 30 in computational time. However, on the other hand some restricting conditions concerning the orbit design and the continuity of the data stream have to be fulfilled to apply the *semi-analytic approach*, while all such phenomena directly related to the space domain such as data gaps are easier to cope with applying a *direct method*.

References

[1] ESA SP-1233(1) (1999): The Four Candidate Earth Explorer Core Missions - Gravity Filed and Steady-State Ocean Circulation Mission. Granada 1999.

[2] ESA (1998): "From Eötvös to Milligal", Final Report of the ESA study 13392/98/NL/GD.

[3] R. Rummel et al., "Spherical harmonic analysis of satellite gradiometry,", Neth. Geod. Comm., no. 39, 1993.

[4] N. Sneeuw, "A semi-analytical approach to gravity field analysis from satellite observations,", Dissertation, DGK, no. 527, 2000.

[5] W.-D. Schuh, "Tailored Numerical Solution Strategies for the Global Determinations of the Earth's Gravity Field," Mitteilungen der geodätischen Institute der TU Graz, no. 81, 1996.

[6] W.-D. Schuh, H. Sünkel, W. Hausleitner, E. Höck, "Refinement of Iterative Procedures for the Reduction of Spaceborne Gravimetry Data," ESA-Project CIGAR IV, WP 4, Final Report, pp. 157-212, 1996.